

UNIVERSALITY OF FIXATION PROBABILITIES IN RANDOMLY STRUCTURED POPULATIONS

BEN ADLAM^{1,2} AND MARTIN A. NOWAK^{1,3,4}

*Program for Evolutionary Dynamics,¹ School of Engineering and Applied Science,²
Department of Mathematics,³ and Department of Organismic and Evolutionary Biology,⁴
Harvard University, Cambridge MA 02138, USA*

ABSTRACT: The stage of evolution is the population of reproducing individuals. The structure of the population is known to affect the dynamics and outcome of evolutionary processes, but analytical results for generic random structures have been lacking. The most general result so far, the isothermal theorem, assumes the propensity for change in each position is exactly the same, but realistic biological structures are always subject to variation and noise. We consider a population of finite size n under constant selection whose structure is given by a wide variety of weighted, directed, random graphs; vertices represent individuals and edges interactions between individuals. By establishing a robustness result for the isothermal theorem and using large deviation estimates to understand the typical structure of random graphs, we prove that for a generalization of the Erdős-Rényi model the fixation probability of an invading mutant is approximately the same as that of a mutant of equal fitness in a well-mixed population with high probability. Simulations of perturbed lattices, small-world networks, and scale-free networks behave similarly. We conjecture that the fixation probability in a well-mixed population, $(1 - r^{-1})/(1 - r^{-n})$, is universal: for many random graph models, the fixation probability approaches the above function uniformly as the graphs become large.

In physics, a system exhibits universality when its macroscopic behavior is independent of the details of its microscopic interactions [13]. Many physical models are conjectured as universal and long programs have been carried out to establish this mathematically [17, 7]. However such universality conjectures have been lacking in biological models.

It is well known that population structure can affect the behavior of evolutionary processes under both constant selection [29, 28, 16, 30, 8, 19], on which we focus here, and frequency dependent selection [40, 25, 37, 11, 9]. However, so far, deterministic and highly organized population structures have received the most attention [24, 10, 22, 14, 23] while some populations are accurately modeled in this way [15, 21, 39, 26, 1],

often a random structure is far more appropriate to describe the irregularity of the real world [18, 42, 5, 34]. Random population structures have been considered numerically, but analytical results have been lacking [30, 37, 4].

The *Moran process* considers a population of n individuals, each of which is either wild-type or mutant with constant fitness 1 or r respectively, undergoing reproduction and death [33]. At each discrete time step an individual is chosen randomly for reproduction proportional to its fitness; another individual is chosen uniformly at random for death and is replaced by a new individual of the same phenotype as the reproducing individual. In the long run, the process has only two possible outcomes: the mutants fix and the wild-type dies out or the reverse. When a single mutant is introduced randomly into a homogenous, wild-type population, we call the probability of the first eventuality *the fixation probability*.

Fixation probabilities are of fundamental interest in evolutionary dynamics [38]. For a well-mixed population as described above, the fixation probability, denoted

$$\rho_{M_n}(r) = \frac{1 - r^{-1}}{1 - r^{-n}}, \quad (1)$$

depends on r and n [31, 35]. Fixation probabilities also depend on population structure [6, 43], which is modeled by running the process on a graph (a collection of n vertices with edges between them) where vertices represent individuals and edges competition between individuals. Population structure forces reproducing individuals to replace only individuals with whom they are in competition, as described by the graph, and thus death is no longer uniformly at random but among only the reproducing individual's neighbors. See the Appendix for details.

With this enrichment of the model, the effects of population structure can be understood. Simple one-rooted population structures are able to repress selection and reduce evolution to a standstill, while intricate, star-like structures can amplify the intensity of selection to all but guarantee the fixation of mutants with arbitrarily slight fitness advantages [30]. The former has been proposed as a model for understanding the necessity of hierarchical lineages of cells to reduce the likelihood of cancer initiation [36]. Some population structures have fixation probabilities which are given exactly by $\rho_{M_n}(r)$ and a fundamental result, called the isothermal theorem (stated precisely in Theorem A.1), gives conditions for this [30]. As a special case of these conditions are all symmetric population structures or graphs with undirected edges. More generally, a graph is called *isothermal* if the sums of the outgoing and ingoing edge weights are the same for all subsets of the graph's vertices. This is our first hint of universality but it was not the first time certain quantities were observed as independent of population structure. Maruyama introduced geographical population structure by separating reproduction, which occurs within sub-populations, and migration, which occurs between sub-populations, and found that the fixation probability was the same as that of a well-mixed population structure [32]. In the framework of evolutionary graph theory, Maruyama's model would correspond to a symmetric graph. In this sense his finding is a special case of the isothermal theorem.

However, the assumptions of the isothermal theorem sit on a knife edge—when any small perturbation is made to the graph, the assumptions no longer hold and the original isothermal theorem is silent. In particular, it cannot be applied to directed, random graphs. We address these shortcomings in Section A, where we strengthen the forward direction of the isothermal theorem by proving a deterministic statement: we weaken the theorem's assumptions to be only approximately true for a graph G and show that the conclusion is still approximately true, that is, the fixation probability of a general graph $\rho_{G_n}(r)$ is approximately equal to $\rho_{M_n}(r)$. We call this the robust isothermal theorem (RIT).

THEOREM (ROBUST ISOTHERMAL THEOREM). *Fix $0 \leq \varepsilon < 1$. Let $G_n = (V_n, W_n)$ be a connected graph. If*

for all nonempty $S \subsetneq V_n$ we have

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| \leq \varepsilon, \quad (2)$$

where $w_O(S)$ and $w_I(S)$ are the sums of the outgoing and ingoing edges respectively, then

$$\sup_{r>0} |\rho_{M_n}(r) - \rho_{G_n}(r)| \leq \varepsilon. \quad (3)$$

This verifies something essential for the process: as in physics, our laws should not depend on arbitrarily small quantities nor make disparate predictions for small perturbations of a system. The RIT generalizes the isothermal theorem in this sense; if an isothermal graph is perturbed with strength ε such that the assumption (2) holds, then its fixation probability is close to that of the original graph (Figure 1). There are many ways of rigorously perturbing a graph, so we do not make a precise definition of perturbation here. All we claim is that any perturbation which changes the assumptions of the RIT continuously can be controlled. The RIT has many useful applications and is our first ingredient to universality.

Robustness is essential for the analysis of random graphs. We say a random graph model exhibits universal Moran-type behavior if its fixation probability behaves like $\rho_{M_n}(r)$ as the graph becomes large. That is, as the graphs become large their macroscopic properties, fixation probabilities, are independent of their microscopic structures, the distributions of individual edges. Mathematically, we ask that the random variable $\sup_{r>0} |\rho_{G_n}(r) - \rho_{M_n}(r)|$ converges in probability to 0, as n goes to infinity. For finite values of n , we can require finer control over this convergence such that

$$\mathbb{P} \left[\sup_{r>0} |\rho_{G_n}(r) - \rho_M(n, r)| \leq \delta(n) \right] = 1 - \varepsilon(n), \quad (4)$$

where the functions $\delta(n) = o(1)$ and $\varepsilon(n) = o(1)$ can be specified. For the generalized Erdős-Rényi model [18] where edges are produced independently with fixed probability p (see Definitions B.4 and D.1) we prove universality. In Sections B and D we analyze the typical behavior of random graphs and show that with very high probability they satisfy the assumptions of the RIT, giving us the paper's main result:

THEOREM. *Let $(G_n)_{n \geq 1}$ be a family of random graphs where the directed edge weights are chosen independently according to some suitable distribution (the outgoing edges may be normalized to sum to 1 or not). Then there are constants $C > 0$ and $c > 0$, not dependent on n , such that the fixation probability of a randomly placed mutant of fitness $r > 0$ satisfies*

$$|\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \quad (5)$$

uniformly in r with probability greater than $1 - \exp(-\nu (\log n)^{1+\xi})$, for some positive constants ξ and ν .

This theorem isolates the typical behavior of the Moran process on these random structures. It can be interpreted as stating that random processes generating population structures where vertices and edges are treated independently and interchangeably will almost always produce graphs with Moran-type behavior. While such processes can generate graphs which do not have Moran-type behavior (for example one-rooted or disconnected graphs), these graphs are generated with very low probability as the size of the graphs becomes large. Moreover, it improves upon diffusion approximation methods by explicitly controlling the error rates [20].

The result holds with high probability but sometimes this probability becomes close to 1 only as the graphs become large. The necessary graph size depends on the distribution that the random graph's edge

weights are drawn from. In particular, it depends inversely on the parameter p from the generalized Erdős-Rényi model, which is the probability that there is an edge of some weight between two directed vertices. The smaller this parameter the more disordered and sparse the random graphs and the less uniform their vertices' temperatures, which all tend to decrease the control over the graph's closeness to isothermality, (2). Regardless, our choice of the parameter $(\log n)^{1+\xi}$ guarantees that the bound [B.54] decays to 0 and that it holds with probability approaching 1 as n becomes large.

We investigated the issues of convergence for small values of n numerically to illustrate our analytical result (Figure 2). For Erdős-Rényi random graphs (see Section B with the distribution chosen as Bernoulli), we generated 10 random graphs according to the procedure outlined in Definition B.4 for fixed values of $0 < p < 1$. On each graph the Moran process was simulated 10^4 times for various values of $0 \leq r \leq 10$ to give the empirical fixation probability, that is, the proportion of times that the mutant fixed in the simulation. Degenerate graphs were not excluded from the simulations but rather than estimating their fixation probabilities, we calculated them exactly, so that 1-rooted graphs were given fixation probability $1/n$ and many-rooted and disconnected graphs were given fixation probability 0. Trivially, such 1-rooted graphs are repressors—that is, the fixation probability of a mutant of fitness $0 < r < 1$ (and a mutant of fitness $r > 1$) is greater than (and less than respectively) the mutant's fixation probability in a well-mixed population—but repressor graphs without these degenerate properties were also observed. As the graphs become larger their fixation probabilities match ρ_{M_n} closely and degeneracy becomes highly improbable as predicted by our result.

In addition to the generalized Erdős-Rényi random graphs, we also considered the Watts-Strogatz model and the Barabási-Albert model. The Watts-Strogatz model [42] produces random graphs with small-world properties, that is, high clustering and short average path length. The model has three inputs: a parameter $0 \leq \beta \leq 1$, the graph size n , and the mean degree $2k$. Typically, the model produces random, undirected graphs, thus, to escape isothermality, it was modified slightly to produce weighted, directed graphs. We do this in the most natural way: we start with a directed $2k$ -regular graph where each node is connected to its $2k$ nearest neighbors if the graph is arranged on a cycle (see Figure 3), and then we rewire each edge to a new vertex chosen uniformly at random with probability β independently. Since the number of edges leaving each vertex is fixed at $2k$, the weight of each edge is exactly $1/(2k)$. Potentially, there can be multiple edges for one vertex to another, which we account for by summing the edge weights. The model may be viewed as an interpolation between an isothermal, $2k$ -regular graph and an Erdős-Rényi graph by the parameter β .

Moran-type behavior was observed in the Watts-Strogatz model for all values of the input parameters we simulated (Figure 3). While mathematical proof of universality in the Watts-Strogatz model is still needed, there is hope that the techniques of this paper may be applied in this situation as the in-degrees of the vertices are concentrated around 1 for graphs with large degree $2k$.

Unlike the Erdős-Rényi and Watts-Strogatz models, scale-free networks are random graphs where the in-degrees of the vertices follow a power law. Normally, scale-free networks are undirected and unweighted. To produce weighted, directed scale-free networks, we modified the preferential attachment algorithm of Barabási-Albert [3]: we start with a connected cycle and then add directed edges of equal weight in sequence to a randomly selected vertex where the destination of each edge is selected proportional to the in-degree of the current vertices.

Surprisingly, even though there is a sense in which vertices are not treated interchangeably in the preferential attachment algorithm, Moran-type behavior was observed in all simulations (Figure 4). This is in contrast with the results in Lieberman *et al.* where they observed some amplification in scale-free networks [30]. The scale-free property is emergent and only becomes apparent as the graph becomes large, thus this increases the running time of the Monte Carlo method for estimating the fixation probability. More simulations are required here for conclusive findings and again there are currently no mathematical results.

In summary, we have generalized the isothermal theorem to make it biologically realistic and to increase its technical applicability. The conclusion of the robust isothermal theorem now depends continuously on its assumptions. With this new tool, we have proved analytically that fixation probabilities in a generalized Erdős-Rényi model converge uniformly to the fixation probability of a well-mixed population. In our proof, we identify the reason for this convergence and bound its rate. Thus, we confirm observations from many simulations and give a method of approximation with a specified error. Furthermore, we conjecture that many random graph models exhibit this universal behavior. However, it is easy to construct simple examples of random graphs which do not, thus it still remains to determine the necessary assumptions on the random graph model for it to exhibit universal behavior.

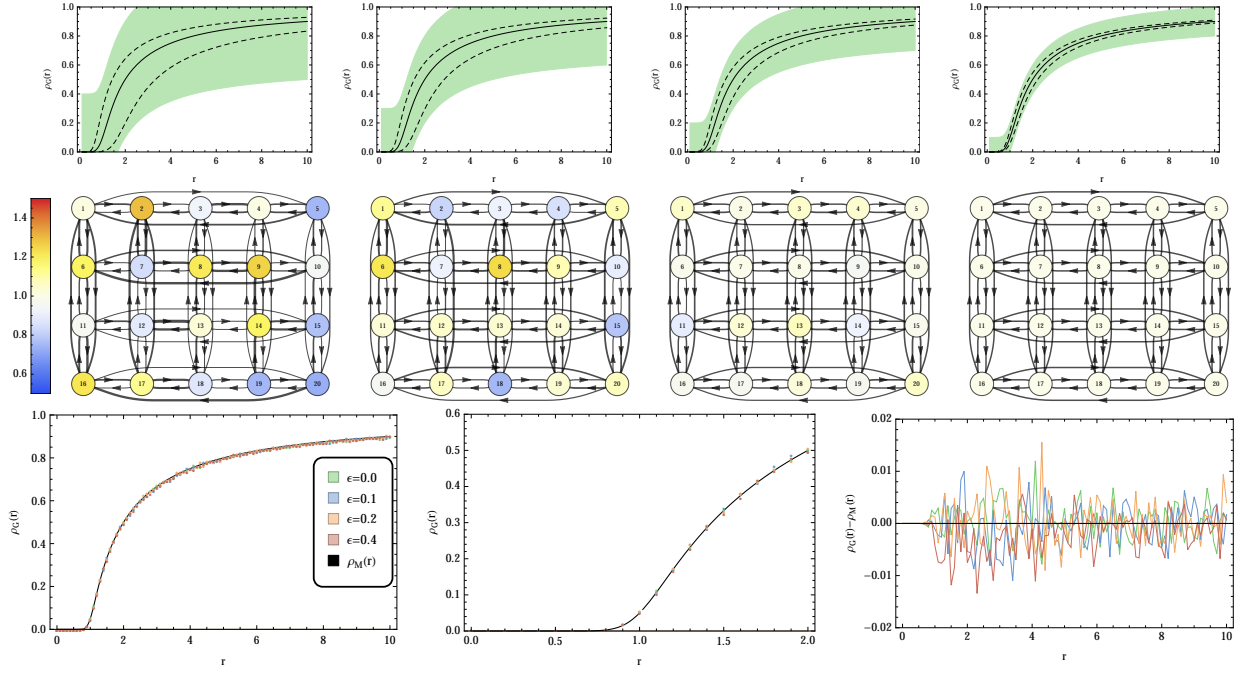


Figure 1: The robust isothermal theorem guarantees that the fixation probability of an approximately isothermal graph lies in the green region and dashed lines indicate the optimal bound. As the perturbation strength decreases through $\varepsilon \in [0.4, 0.1]$ and the graph approaches isothermality, the bound improves and converges uniformly to the solid black line, ρ_{M_n} . The figures of square lattices show how random perturbations shift the graphs from isothermality, as the perturbation strength decreases from left to right; we draw each graph with the directed edges' thickness proportional to their weight and the vertices' color given by the sum of the weights of edges pointing to them. In the bottom row empirical estimates of the fixation probabilities (small circles) are plotted against the values predicted by ρ_{M_n} (solid lines) and, despite the perturbations to the graphs, their fixation probabilities lie close to ρ_{M_n} .

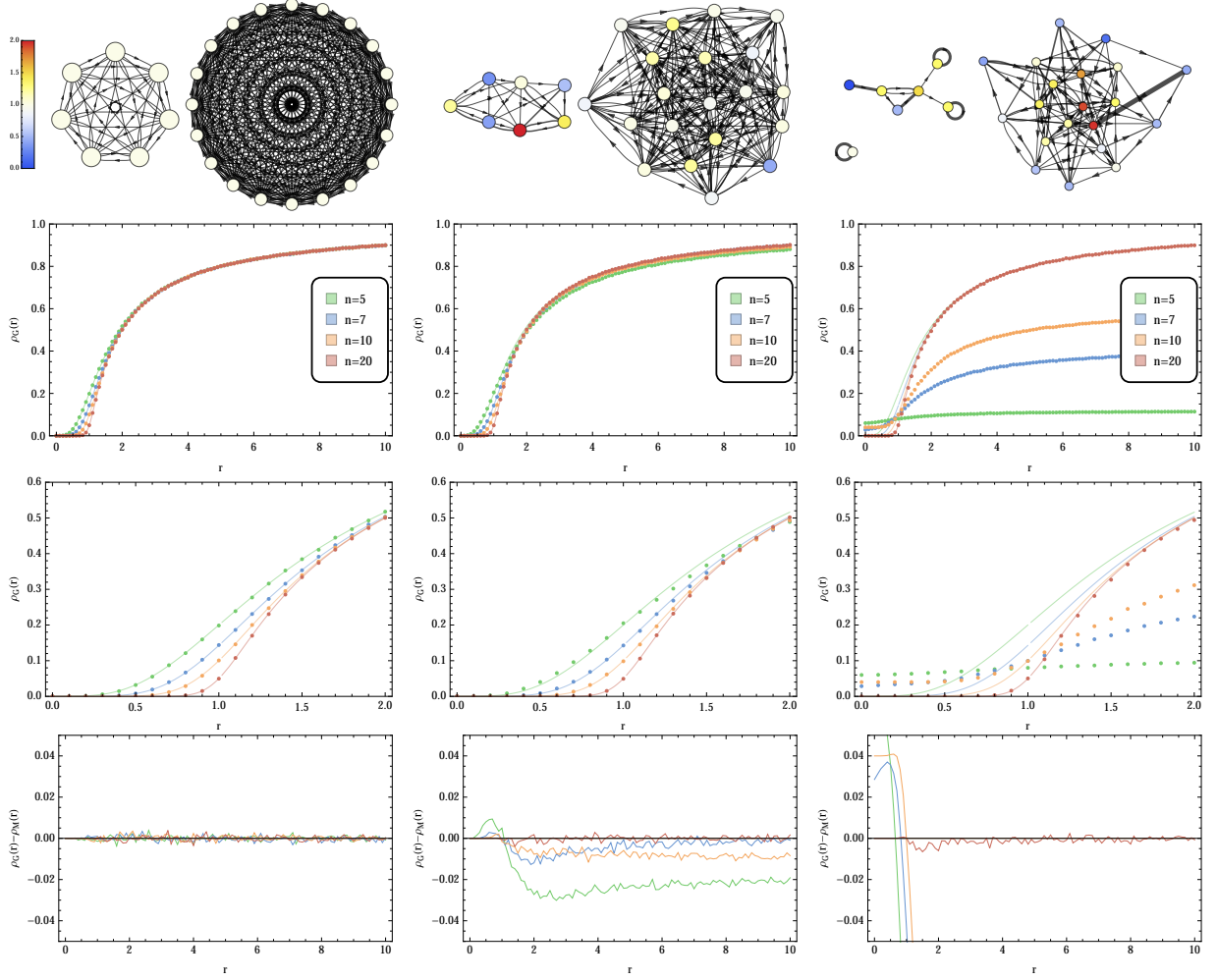


Figure 2: The fixation probability of the generalized Erdős-Rényi random graphs converge uniformly to ρ_{M_n} . The three columns from left to right correspond to Erdős-Rényi random graphs with decreasing connection probabilities $p = 1$, $p = 0.6$, and $p = 0.3$. The representative random graphs in the top row show both the increasing sparsity and disorder as p decreases and the elimination of degeneracy (rootedness and disconnectedness) and the increasing uniformity of temperature as the graph sizes increase. In the middle two rows empirical estimates of the fixation probabilities (small circles) are plotted against the values predicted by ρ_{M_n} (solid lines). When $p = 1$ the graphs are isothermal and thus correspond exactly to their predicted values which can be seen even more clearly in the bottom row, where the difference of the empirical fixation probabilities and their predicted values display as stochastic fluctuations about 0. For $p = 0.6$ and $p = 0.3$, the convergence of the empirical values to ρ_{M_n} as the graphs increases in size is apparent. Smaller graphs are typically repressors as illustrated by the clear sign change at $r = 1$ in the difference of empirical and predicted values, whereas larger graphs fluctuate about 0. Moreover, the convergence is patently slower in n for smaller values of p .

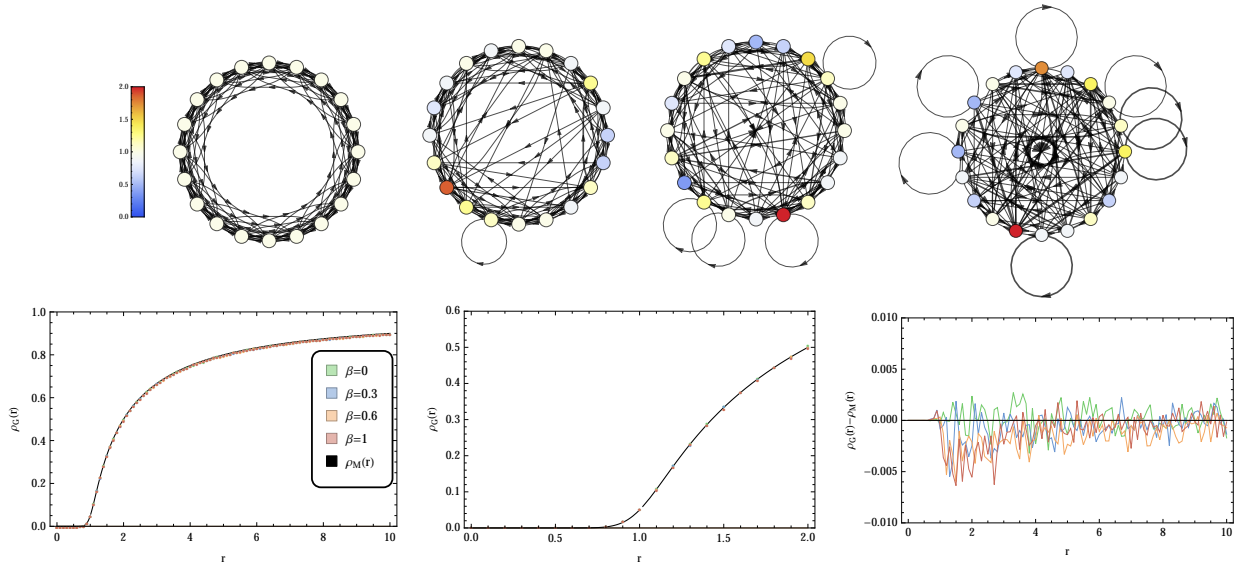


Figure 3: Small-world networks also show universal behavior. Representative Watts-Strogatz random graphs display increasing disorder as the rewiring probability β increases from 0 to 1, which may be viewed as an interpolation between an isothermal graph to an Erdős-Rényi random graph. For all values of β the correspondence to ρ_{M_n} is striking but mathematical proof is lacking.

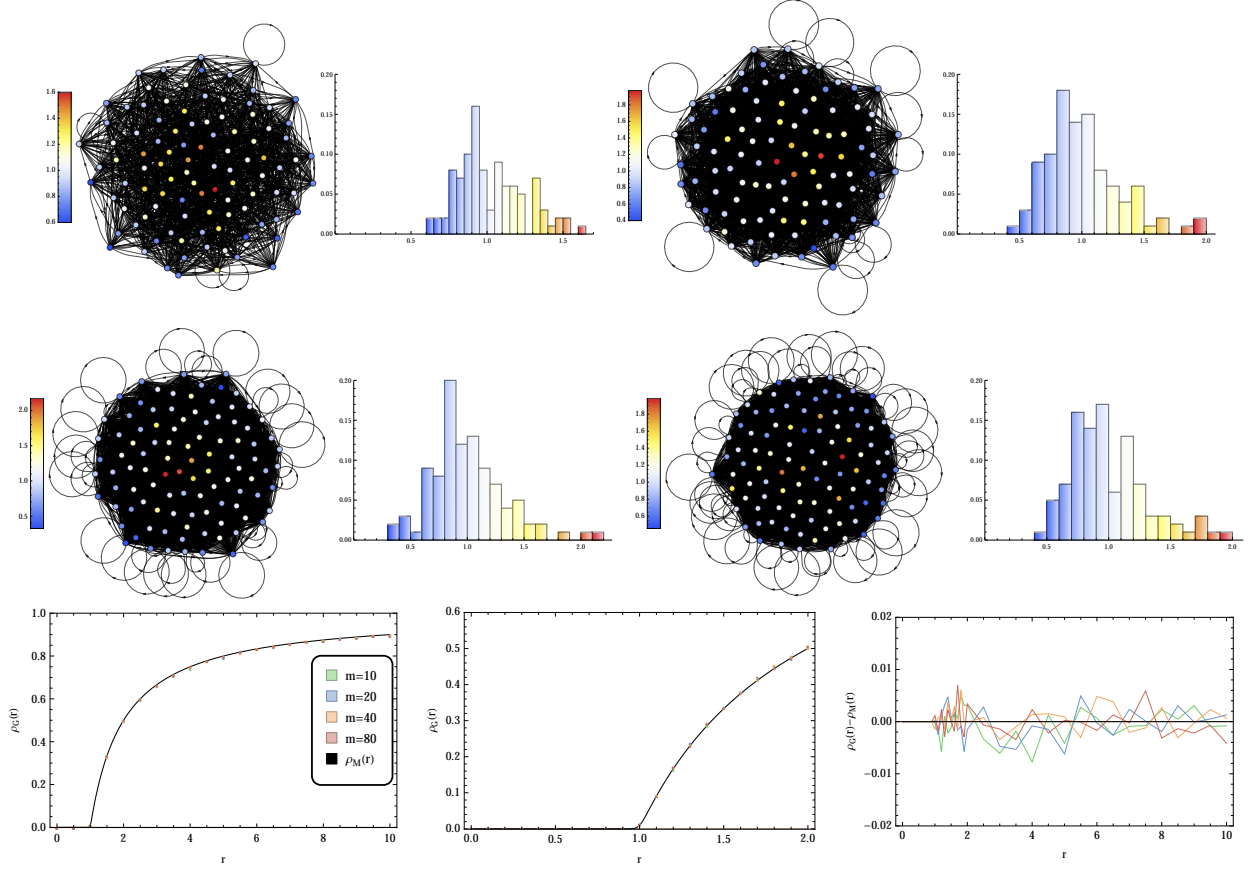


Figure 4: Simulations on graphs generated by preferential attachment yield fixation probabilities close to ρ_{M_n} . Several scale-free networks with varying out degrees, $m = 10$, $m = 20$, $m = 40$, and $m = 80$, were generated using a preferential attachment algorithm. Histograms of the sum of the weights of edges pointing to each vertex are plotted next to each graph, however, the small graphs size limits the resemblance to a power law. Given the comparatively large size of the graphs, only a restricted number of simulations were performed, but the simulations corresponded to ρ_{M_n} without a tendency to amplify or repress. More extensive work is required.

APPENDIX A. THE ROBUST ISOTHERMAL THEOREM

The Moran process on graphs is the standard model for population structure in evolutionary dynamics [30, 35]. The process is defined for a directed, weighted graph, $G \equiv G_n = (V_n, W_n)$, where $V \equiv V_n := \llbracket n \rrbracket$ and $W \equiv W_n := [w_{ij}]$ is a stochastic matrix of edge weights. The process X_t is Markovian with state space 2^V , where each X_t is the subset of V consisting of all the vertices occupied by mutants at time t . At time 0 a mutant is placed at one of the vertices uniformly at random or formally,

$$\mathbb{P}[X_0 = S] = \begin{cases} n^{-1} & \text{if } |S| = 1 \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A.1})$$

Then at each subsequent time step exactly one vertex is chosen randomly, proportional to its fitness, for reproduction: so the probability of choosing a particular wild type vertex is $1/(n - |X_t| + r|X_t|)$ and the probability of choosing a particular mutant vertex is $r/(n - |X_t| + r|X_t|)$. An edge originating from the chosen vertex is then selected randomly with probability equal to its edge weight, which is well defined since W is stochastic, and the vertex at the destination of the edge takes on the type of the vertex at the origin of the edge.

Typically, there are exactly two absorbing states, $X_t = \emptyset$ and $X_t = V$, corresponding to the wild type fixing in the population and the mutant fixing in the population respectively. Thus, almost surely, one of these two absorbing states is reached in finite time. The probability that the process reaches V and not \emptyset is called the *fixation probability* and for a graph G we denote its fixation probability for a mutant of fitness $r > 0$ by

$$\rho_G(r) := \mathbb{P}[X_t = V \text{ for some } t \geq 0]. \quad (\text{A.2})$$

A fundamental point of comparison is the fixation probability ρ_{M_n} for a *well-mixed* population structure, where the graph structure is given by

$$w_{ij} := \begin{cases} (n-1)^{-1} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (\text{A.3})$$

and M stands for “Moran” or “mixed.” An easy calculation using recurrence equations shows that

$$\rho_{M_n}(r) = \frac{1 - r^{-1}}{1 - r^{-n}}. \quad (\text{A.4})$$

Graphs with exactly the same fixation probability as ρ_{M_n} are classified by the isothermal theorem, which gives sufficient conditions for a general graph G to have the same fixation probability as ρ_{M_n} [30].

In this section we derive a generalization of the isothermal theorem and throughout we shall require that the matrix W is stochastic—that is, the row sums are all equal to 1:

$$\sum_{j=1}^n w_{ij} = 1, \quad (\text{A.5})$$

for all $i \in V$. Any graph with nonnegative edge weights can be normalized to produce a graph with a stochastic W , so long as each row has a nonzero entry, without changing the behavior of the process as defined above. A graph G is called isothermal if all the column sums of W are identical—that is, $W_1 = \dots = W_n = 1$, where

$$W_j := \sum_{i=1}^n w_{ij}, \quad (\text{A.6})$$

or equivalently, W is doubly stochastic.

For all $S \subseteq V$, define

$$w_O(S) := \sum_{i \in S} \sum_{j \notin S} w_{ij} \quad \text{and} \quad w_I(S) := \sum_{i \notin S} \sum_{j \in S} w_{ij} \quad (\text{A.7})$$

as the sum of the edge weights leaving S and entering S respectively. Then an easy calculation shows that a graph is isothermal if and only if

$$w_O(S) = w_I(S) \quad (\text{A.8})$$

for all $\emptyset \neq S \subsetneq V$. The later condition and its equivalence to isothermality is at the core of the proof of the isothermal theorem. The term “isothermal” originates from an interpretation of the sum of the ingoing edge weights as temperature, with “hotter” vertices changing more frequently in the Moran process. Thus a graph satisfying (A.8) is isothermal because the ingoing and outgoing temperatures are equal and all subsets S are in “thermal equilibrium.” We now restate the forward direction of the original isothermal theorem.

THEOREM A.1 (ISOTHERMAL THEOREM). *Suppose that a graph G is isothermal, then the fixation probability of a randomly placed mutant of fitness r is equal to $\rho_{M_n}(r)$.*

We ask, can we relax the assumptions of Theorem A.1? That is, perhaps an approximate result can be obtained for W that are only approximately doubly stochastic in the following sense:

$$|W_j - 1| \leq \varepsilon \quad (\text{A.9})$$

for all $j \in V$ and some small quantity ε . However, the example

$$W = \begin{bmatrix} 0 & 1 - \varepsilon & \varepsilon & 0 \\ 1 - \varepsilon & 0 & 0 & \varepsilon \\ \varepsilon^2 & 0 & 0 & 1 - \varepsilon^2 \\ 0 & \varepsilon^2 & 1 - \varepsilon^2 & 0 \end{bmatrix} \quad (\text{A.10})$$

shows we cannot, since as $\varepsilon \rightarrow 0$

$$\frac{w_O(\{1, 2\})}{w_I(\{1, 2\})} = \frac{2\varepsilon}{2\varepsilon^2} = \varepsilon^{-1} \rightarrow \infty. \quad (\text{A.11})$$

That is, W is approximately doubly stochastic, but the ratio of the outgoing and ingoing edge weights is unbounded for some subset S . Thus, we need stronger assumptions for our theorem which we state now.

THEOREM A.2 (ROBUST ISOTHERMAL THEOREM). *Fix $0 \leq \varepsilon < 1$. Let $G_n = (V_n, W_n)$ be a connected graph. If for all nonempty $S \subsetneq V_n$ we have*

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| \leq \varepsilon, \quad (\text{A.12})$$

then

$$\sup_{r > 0} |\rho_{M_n}(r) - \rho_{G_n}(r)| \leq \varepsilon. \quad (\text{A.13})$$

PROOF. To briefly outline the proof, we begin by projecting the process from X_t to $|X_t|$. Next we consider the ratio of the probability of increasing the number of mutants to the probability of decreasing the number of mutants. By bounding this ratio we can use a coupling argument to establish that the fixation probability of the process is close to ρ_{M_n} . Finally, we use the mean value theorem and smoothness properties of ρ_{M_n} to simplify our bound and obtain the result.

Just as in the proof of the original isothermal theorem, we make the projection of the state space of all subsets of V , which records exactly which vertices are mutants, to the simpler state space $\{0, 1, \dots, n\}$, which records only the number of mutants. The problem with making this projection in general is that the transition probabilities from one subset to another can depend on the structure of a subset not merely the number of mutants. However, it is clear that the only quantities which affect the fixation probability are the ratios of the probability of increasing the number of mutants to the probability of decreasing the number of mutants in a particular state S .

Define $p_+(S)$ and $p_-(S)$ as the probability that the number of mutants in a population increases and decreases by one respectively. Thus

$$p_+(S) = \frac{w_O(S)r}{w_O(S)r + w_I(S)} \quad \text{and} \quad p_-(S) = \frac{w_I(S)}{w_O(S)r + w_I(S)}, \quad (\text{A.14})$$

which gives, when the two equations are divided,

$$\frac{p_+(S)}{p_-(S)} = r \frac{w_O(S)}{w_I(S)}. \quad (\text{A.15})$$

By assumption (A.12),

$$r(1 - \varepsilon) \leq \frac{p_+(S)}{p_-(S)} \leq r(1 + \varepsilon). \quad (\text{A.16})$$

This states that the ratio of the probabilities of increasing to decreasing the number of mutants in any state S is approximately proportional to r .

If for some graph $G' = (V', W')$ we have $p_+(S)/p_-(S) = r(1 \pm \varepsilon)$ for all $S \subseteq V'$, then by the standard result for fixation probabilities in birth-death processes, its fixation probability is given by

$$\rho_{G'}(r) = \rho_{M_n}(r \pm r\varepsilon) = \frac{1 - (r \pm \varepsilon r)^{-1}}{1 - (r \pm \varepsilon r)^{-n}}. \quad (\text{A.17})$$

From (A.16) and (A.17) we would like to conclude that

$$\rho_{M_n}(r - r\varepsilon) = \frac{1 - (r - \varepsilon r)^{-1}}{1 - (r - \varepsilon r)^{-n}} \leq \rho_G(r) \leq \frac{1 - (r + \varepsilon r)^{-1}}{1 - (r + \varepsilon r)^{-n}} = \rho_{M_n}(r + r\varepsilon). \quad (\text{A.18})$$

The upper bound is given by taking the maximum allowed value for the probability of increasing the number of mutants relative to the probability of decreasing the number of mutants. For the lower bound we use the opposite.

This intuitive result can be proved with a coupling argument. We can couple the Moran process X_t of a mutant of fitness r on G with another process Y defined as follows: Y has state space $\{0, \dots, n\}$ (with 0 and n absorbing) and Y starts at 1. We couple Y to X_t as follows:

- (i) if $|X_t|$ decreases by 1, then Y must also decrease by 1;
- (ii) if $|X_t|$ increases by 1, then independently Y increases by 1 with probability

$$\frac{P_+(S) + P_-(S)}{P_+(S)} \frac{r(1 - \varepsilon)}{1 + r(1 - \varepsilon)} \quad (\text{A.19})$$

(which is less than or equal to 1 by assumption (A.16)), else Y decreases by 1;

(iii) otherwise Y remains constant.

Note that marginally Y is a simple random walk on $\{0, \dots, n\}$ with forward bias $r(1 - \varepsilon)$ and thus the probability that Y reaches n before it reaches 0 is given by

$$\frac{1 - (r - \varepsilon r)^{-1}}{1 - (r - \varepsilon r)^{-n}}. \quad (\text{A.20})$$

However, because the processes are coupled we have $Y \leq |X_t|$ and thus if $Y = n$, then the mutant has fixed in the process X_t . Equation (A.20) immediately implies the lower bound in (A.18). A similar coupling yields the upper bound. Thus

$$\rho_{M_n}(r - r\varepsilon) - \rho_{M_n}(r) \leq \rho_G(r) - \rho_{M_n}(r) \leq \rho_{M_n}(r + r\varepsilon) - \rho_{M_n}(r). \quad (\text{A.21})$$

By the mean value theorem,

$$\frac{\rho_{M_n}(r + r\varepsilon) - \rho_{M_n}(r)}{\varepsilon} \leq \frac{r\varepsilon}{\varepsilon} \sup_{r \leq x \leq r+r\varepsilon} |\rho'_{M_n}(x)| = r \sup_{r \leq x \leq r+r\varepsilon} |\rho'_{M_n}(x)| \quad (\text{A.22})$$

and

$$\frac{\rho_{M_n}(r) - \rho_{M_n}(r - r\varepsilon)}{\varepsilon} \geq -r \sup_{r-r\varepsilon \leq x \leq r} |\rho'_{M_n}(x)|. \quad (\text{A.23})$$

Thus, it is sufficient to show for all $r > 0$

$$\sup_{n \geq 2} |\rho'_{M_n}(r)| \leq r^{-1}. \quad (\text{A.24})$$

We note that this is not an optimal bound, however, it suffices for our applications. Calculating, one finds

$$\rho'_{M_n}(r) = \frac{r^{n-2}(r^n - nr + n - 1)}{(r^n - 1)^2}. \quad (\text{A.25})$$

First, when $r \geq 1$ we prove the stronger claim

$$\frac{r^{n-2}(r^n - nr + n - 1)}{(r^n - 1)^2} \leq r^{-2}, \quad (\text{A.26})$$

by noting the above is equivalent to

$$(r - 1) \left(nr^n - \sum_{k=0}^{n-1} r^k \right) \geq 0, \quad (\text{A.27})$$

which is true since $r \geq 1$. Similarly, one can prove

$$\frac{r^{n-2}(r^n - nr + n - 1)}{(r^n - 1)^2} < 1 \quad (\text{A.28})$$

when $r < 1$. Equations (A.26) and (A.28) imply $\sup_{n \geq 2} |\rho'_{M_n}(r)| \leq r^{-1}$.

Therefore, we may conclude

$$\sup_{r > 0} |\rho_G(r) - \rho_{M_n}(r)| \leq \varepsilon, \quad (\text{A.29})$$

which completes the proof. ■

Note that one can sometimes do slightly better (depending on the relative sizes of ε and n^{-1}), by showing

$$\sup_{r \leq 1} |\rho_{G_n}(r) - \rho_{M_n}(r)| < \frac{2}{n} \quad (\text{A.30})$$

for ε small enough, using the triangle inequality and the facts that ρ_{M_n} is increasing and continuous, but again this is not important for our applications.

Theorem A.2 is actually slightly stronger than stated, and thus we can draw a slightly stronger conclusion in Theorem B.1. We may conclude that the fixation probability of a mutant of fitness r originating at a particular vertex i [2] satisfies the bound in (A.13), for exactly the same reason as in the proof of the original isothermal theorem—the bound (A.12) is for all subsets S . Therefore, *a fortiori*, a mutant can be started with any probability vector on the vertices (not merely uniform) and its fixation probability will still satisfy (A.13). This observation is borne through simulations too (Figure 5).

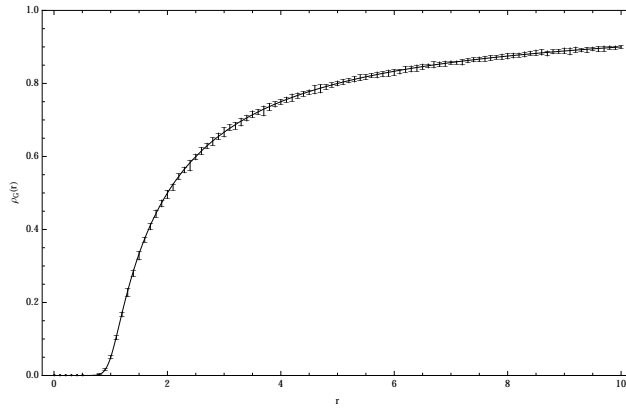


Figure 5: Fixation probability does not depend on starting location. We conducted trials where the fixation probability of a mutant of fitness r starting at vertex i was estimated with the Monte Carlo method of 10^4 samples for several values of $0 \leq r \leq 10$ on a Erdős-Rényi random graph. The fixation probability was similar regardless of starting vertex, and in particular, showed no correlation with vertex temperature. We plot the Moran fixation probability (A.4) and use the error bars to illustrate the minimum and maximum empirical fixation probabilities obtained from starting at any particular vertex.

APPENDIX B. EVOLUTION ON RANDOM GRAPHS

In this section we prove Theorem B.1.

THEOREM B.1. *Let $(G_n)_{n \geq 1}$ be a family of random graphs as in Definition B.4. Then there are constants $C > 0$ and $c > 0$, not dependent on n , such that the fixation probability of a randomly placed mutant of fitness $r > 0$ satisfies*

$$|\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \quad (\text{B.1})$$

uniformly in r with probability greater than

$$1 - \exp\left(-\nu (\log n)^{1+\xi}\right), \quad (\text{B.2})$$

for positive constants ξ and ν .

To do this we need to apply Theorem A.2 to our random graphs by showing that its assumptions hold with high probability. We do so in several steps. First, we define precisely generalized Erdős-Rényi random graphs in Definition B.4 and outline the necessary assumptions on the distribution of the edge weights. After reviewing some notation, we introduce an event Ω , on which the graphs are well behaved, and show that Ω has high probability in Lemma B.6. Then the general idea is to use large deviation estimates and concentration inequalities to show that with high probability the quantity (A.12) can be controlled. We bound both the numerator (Lemma B.7) and denominator (Lemma B.8) of

$$\frac{|w_O(S) - w_I(S)|}{|w_I(S)|} = \left| \frac{w_O(S)}{w_I(S)} - 1 \right| \quad (\text{B.3})$$

for all S , then we put everything together to prove Theorem B.1.

REMARK B.2 (Notation). We use the large constant $C > 0$ and the small constant $c > 0$, which do not depend on the size of the graph n but can depend on the distribution as outlined in Definition B.4. We allow the constant C to increase or the constant c to decrease from line to line without noting it or introducing a new notation, sometimes we even absorb other constants such as p , p' , and μ_1 without noting it; as is clear from the proof, this only happens a finite number of times, and thus we end with constants $C > 0$, $c > 0$, $\nu > 0$, and $\xi > 0$.

We also make use of standard order notation for functions, $o(\cdot)$, $\mathcal{O}(\cdot)$, and $\cdot \gg \cdot$, all of which are used with respect to n . Moreover, in some sums it is useful to exclude particular summands, e.g.

$$\sum_{\substack{1 \leq j \leq n \\ j \notin S}} \cdot \equiv \sum_j^{(S)} \cdot \quad (\text{B.4})$$

for $S \subset V$. We abbreviate $(\{i\})$ and $(\{i, k\})$ as (i) and (i, k) .

REMARK B.3 (High probability events). We say that an n -dependent event E holds with high probability if, for constants $\xi > 0$ and $\nu > 0$ which do not depend on n ,

$$\mathbb{P}[E^c] \leq e^{-\nu(\log n)^{1+\xi}} \quad (\text{B.5})$$

for $n \geq n_0(\nu, \xi)$. Moreover, we say an event E has high probability on another event E_0 if

$$\mathbb{P}[E_0 \cap E^c] \leq e^{-\nu(\log n)^{1+\xi}} \quad (\text{B.6})$$

In particular, this has the property that the intersection of polynomially many (in n , say Kn^K for some constant $K > 0$) events of high probability is also an event of high probability: by the union bound,

$$\mathbb{P}\left[\left(\bigcap_{i=1}^{Kn^K} E_i\right)^c\right] = \mathbb{P}\left[\bigcup_{i=1}^{Kn^K} E_i^c\right] \leq Kn^K \cdot e^{-\nu(\log n)^{1+\xi}} = Ke^{K \log n - \nu(\log n)^{1+\xi}} \leq e^{-\nu(\log n)^{1+\xi}}, \quad (\text{B.7})$$

with a possible increase in $n_0(\nu, \xi)$ and a decrease in the constants ν and ξ .

B.1. Proof of Theorem B.1. Following the Erdős-Rényi model, we produce a weighted, directed graph as follows: Consider an $n \times n$ matrix $X = [x_{ij}]$ with zero for its diagonal entries and independent, identically distributed, nonnegative random variables for its off-diagonal entries. We now want to define a random, stochastic matrix W of weights. The natural definition for $W = [w_{ij}]$ is

$$w_{ij} := \frac{x_{ij}}{\sum_{k=1}^n x_{ik}}, \quad (\text{B.8})$$

which is defined when at least one of the x_{i1}, \dots, x_{in} is nonzero; this happens almost surely in the limit as $n \rightarrow \infty$, when $\mathbb{P}[x_{ij} > 0] = p > 0$ is a constant:

$$\mathbb{P}[x_{i1} = x_{i2} = \dots = x_{in} = 0] = (1 - p)^{n-1} \quad (\text{B.9})$$

and by the union bound

$$\mathbb{P}\left[\bigvee_{i=1}^n x_{i1} = x_{i2} = \dots = x_{in} = 0\right] \leq \sum_{i=1}^n \mathbb{P}[x_{i1} = x_{i2} = \dots = x_{in} = 0] = n(1 - p)^{n-1} \rightarrow 0. \quad (\text{B.10})$$

However, the question is how to technically deal with the unusual event that all the entries of a row of X are zero, as there are several options. We make the following choice: for $1 \leq i \leq n$

$$w_{ii} := \begin{cases} 1 & \text{if } x_{i1} = x_{i2} = \dots = x_{in} = 0 \\ 0 & \text{otherwise} \end{cases}, \quad (\text{B.11})$$

and for all $1 \leq i, j \leq n$ and $i \neq j$

$$w_{ij} := \begin{cases} \frac{x_{ij}}{\sum_{k=1}^n x_{ik}} & \text{if } x_{ij} > 0 \\ 0 & \text{if } x_{ij} = 0 \end{cases}. \quad (\text{B.12})$$

This definition aligns with the definition in (B.8) with probability greater than $1 - n(1 - p)^n$. Moreover, this definition has the advantage that the events that any non-loop edge weight is 0 are independent.

DEFINITION B.4 (GENERALIZED ERDŐS-RÉNYI RANDOM GRAPHS). Let μ be a nonnegative distribution (not depending on n) with subexponential decay such that if $X \sim \mu$

$$\mathbb{P}[X > 0] = p > 0 \text{ and } \mathbb{P}[X \geq x] \leq Ce^{-x^c} \quad (\text{B.13})$$

for some positive constants p , c , and C and all $x > 0$. We denote the mean and standard deviation of X by μ_1 and σ respectively. We generate a family of random graphs $G_n = (V_n, W_n)$ from μ by defining the weight matrices W_n according to (B.11) and (B.12), where x_{ij} are independent and distributed according to μ for $i \neq j$.

The subexponential decay is necessary to control the fluctuation of the graph's edge weights and imposes a bounded increase on the moments of μ . Let $X \sim \mu$, then simple calculations show,

$$\mu_k := \mathbb{E}X^k \leq Ck \int_0^\infty x^{k-1} \mathbb{P}[X \geq x] dx \leq Ck \int_0^\infty x^{k-1} e^{-x^c} dx = \frac{C}{c} k \Gamma\left(\frac{1+k}{c}\right) \leq (Ck)^{Ck}, \quad (\text{B.14})$$

where the constant $C > 0$ depends on the constants in (B.13). Many distributions satisfy the subexponential assumption (B.13), for example any compactly supported distribution, the Gamma distribution, and the absolute value of a Gaussian distribution.

We now use the subexponential decay assumption to understand the typical behavior of the random variables x_{ij} .

DEFINITION B.5 (GOOD EVENTS Ω). *Let Ω be an n -dependent event such that the following hold:*

$$\Omega := \bigcap_{i=1}^n \left(\{x_{ii} = 0\} \cap \left\{ \left| \sum_j^{(i)} (x_{ij} - \mathbb{E}x_{ij}) \right| \leq \sigma (\log n)^{C+C\xi} \sqrt{n} \right\} \right) \quad (\text{B.15})$$

$$\cap \bigcap_{i,j=1}^n \left\{ x_{ij} \leq C (\log n)^C \right\} \cap \{G_n \text{ is connected}\}. \quad (\text{B.16})$$

The conditions on Ω have natural interpretations. The first condition specifies that the normalization procedure outlined above has worked as intended and that we are not in the atypical case where the graph has a self-loop. The second condition specifies that the sums of n of the x_{ij} s are close to their expectation $(n-1)\mu_1$ and that they fluctuate about this value on the order of \sqrt{n} as predicted by the central limit theorem. The third condition says that none of the x_{ij} are too large and that typically they will all be less than $C (\log n)^C$. The last condition is self-explanatory, as the Moran process is not guaranteed to terminate on disconnected graphs.

LEMMA B.6. *The event Ω holds with high probability.*

PROOF. By Remark B.3, it suffices to show that each conjunct holds with high probability as there are only polynomially many choices for i and j . First fix i . By assumption (B.13) and the fact that $x_{ii} \neq 0$ only if $x_{ij} = 0$ for all $j \neq i$,

$$\mathbb{P}\{x_{ii} \neq 0\} = \mathbb{P}[x_{ij} = 0 \text{ for all } j \neq i] \leq (1-p)^{n-1} = e^{\log(1-p)(n-1)} \leq e^{-\nu(\log n)^{1+\xi}}, \quad (\text{B.17})$$

since $0 < p < 1$ and $n-1 \gg (\log n)^{1+\xi}$.

Now using the large deviation result, Lemma E.4, with $a_j = x_{ij} - \mathbb{E}x_{ij}$ and $A_j = 1$, we may verify the moment assumption (E.7): clearly $\mathbb{E}(x_{ij} - \mathbb{E}x_{ij}) = 0$ and $\mathbb{E}(x_{ij} - \mathbb{E}x_{ij})^2 = \sigma^2$, then

$$\mathbb{E}|x_{ij} - \mathbb{E}x_{ij}|^k \leq (Ck)^{Ck} \quad (\text{B.18})$$

by Equation (B.14). Thus we get

$$\mathbb{P}\left[\left|\sum_{i=1}^n a_i A_i\right| \geq \sigma (\log n)^{C+C\xi} \sqrt{n}\right] \leq e^{-\nu(\log n)^{1+\xi}}. \quad (\text{B.19})$$

Now fix j too. Next, we use the subexponential decay assumption (B.13) with $x = C (\log n)^C$ for $C = 1 + c^{-1}$ to get

$$\mathbb{P}\left[x_{ij} > C (\log n)^{c^{-1}+1}\right] \leq C \exp\left(-\left(C (\log n)^{c^{-1}+1}\right)^c\right) \leq C e^{-C^c (\log n)^{1+c}} \leq e^{-\nu(\log n)^{1+\xi}}. \quad (\text{B.20})$$

Thus, $x_{ij} > C (\log n)^C$ holds with high probability since $c > 0$ and $C^c > 0$.

Finally, we show that the graph G is connected with high probability, i.e. that with high probability, the graph cannot be partitioned into two disjoint sets where there are no edges going from one subset to another. This follows from an argument similar to that contained in the proof of Lemma B.8 but without the assumption that we are on the event Ω as we do not need a lower bound on the weights only that edges exist which they do with probability at least p . \blacksquare

Note that by definition W is stochastic. Define the sum of the j th column as

$$W_j := \sum_{i=1}^n w_{ij}. \quad (\text{B.21})$$

Note that while the family W_j is not independent, by symmetry, they are identically distributed. Hence

$$\mathbb{E}W_1 = \frac{1}{n} \sum_{j=1}^n \mathbb{E}W_j = \frac{1}{n} \mathbb{E} \sum_{j=1}^n \sum_{i=1}^n w_{ij} = 1. \quad (\text{B.22})$$

This tells us that in expectation W is doubly stochastic. The next lemma shows that with very high probability it is almost $n^{-1/2}$ close to being doubly stochastic, which is exactly the order of fluctuation we expect by the central limit theorem. The assumptions on the distribution μ and the event Ω guarantee that we can prove that the sum's fluctuations are of this order.

The idea of the proof is that for fixed j , the w_{ij} are independent random variables and thus we can apply a LDE to bound the fluctuations of their sum. There are complications due to the normalization required by Definition B.4 but on Ω these can be overcome by relating the sum W_j to a simpler sum that may be controlled with Lemma E.4.

LEMMA B.7. *On Ω , there are positive constants $c \equiv c_\mu$ and $C \equiv C_\mu$, not dependent on n , such that the following inequalities hold*

$$|W_j - 1| \leq \frac{C(\log n)^{C+C\xi}}{\sqrt{n}} \quad (\text{B.23})$$

for all $j \in V$, with probability at least

$$1 - e^{-\nu(\log n)^{1+\xi}}. \quad (\text{B.24})$$

PROOF. Fix j . First we use the fact that $w_{ii} = x_{ii} = 0$ for $1 \leq i \leq n$ on Ω to see

$$W_j - 1 = \sum_i (w_{ij} - \mathbb{E}w_{ij}) = \sum_i^{(j)} \left(w_{ij} - \frac{1}{n-1} \right) + \mathcal{O}(n^2(1-p)^{-n+1}). \quad (\text{B.25})$$

By the definition of w_{ij} , the above is equal to

$$\sum_i^{(j)} \left(\frac{x_{ij}}{\sum_k^{(i)} x_{ik}} - \frac{1}{n-1} \right) + \mathcal{O}(c_0^{-n}) = \sum_i^{(j)} \left(\frac{x_{ij} - \frac{1}{n-1} \sum_k^{(i)} x_{ik}}{\sum_k^{(i)} x_{ik}} \right) + \mathcal{O}(c_0^{-n}), \quad (\text{B.26})$$

where $c_0 < 1$ is not dependent on n . Next, using the fact that on Ω

$$\frac{1}{n-1} \left| \sum_k^{(i)} (x_{ik} - \mathbb{E}x_{ik}) \right| \leq C\sigma(\log n)^{C+C\xi} \frac{1}{\sqrt{n}} \quad (\text{B.27})$$

for all $1 \leq i \leq n$, we replace the average in the numerator of (B.26) with its expectation to find it equal to

$$\sum_i^{(j)} \left(\frac{x_{ij} - \mathbb{E}x_{ij}}{\sum_k^{(i)} x_{ik}} + \frac{C\sigma(\log n)^{C+C\xi}}{\sqrt{n} \sum_k^{(i)} x_{ik}} \right) + \mathcal{O}(c_0^{-n}). \quad (\text{B.28})$$

Using (B.27) again, it is easy to see

$$\sum_k^{(i)} x_{ik} \geq (n-1) \mathbb{E} x_{ij} - C \sigma (\log n)^{C+C\xi} \sqrt{n}, \quad (\text{B.29})$$

which gives an upper bound on the error term in (B.28) and we find the equation equal to

$$\sum_i^{(j)} \frac{x_{ij} - \mathbb{E} x_{ij}}{\sum_k^{(i)} x_{ik}} + \mathcal{O} \left(\frac{(\log n)^{2C+2C\xi}}{\sqrt{n}} \right) + \mathcal{O}(c_0^{-n}). \quad (\text{B.30})$$

Next we compare these two expressions to find that the absolute value their difference can be expressed as

$$\left| \frac{x_{ij} - \mathbb{E} x_{ij}}{\sum_k^{(i)} x_{ik}} - \frac{x_{ij} - \mathbb{E} x_{ij}}{\sum_k^{(i,j)} x_{ik}} \right| = \frac{|x_{ij}|^2}{\left| \sum_k^{(i)} x_{ik} \cdot \left(\sum_k^{(i)} x_{ik} - x_{ij} \right) \right|}. \quad (\text{B.31})$$

However, using that on Ω , for all $1 \leq i, j \leq n$, we have $x_{ij} \leq C (\log n)^C$ and using (B.27) as before, we may show the difference is bounded by

$$\mathcal{O} \left(\frac{(\log n)^{4C+2C\xi}}{n^2} \right). \quad (\text{B.32})$$

We can then sum over these errors—one for each summand—to get a total error of $\mathcal{O} \left((\log n)^{4C+2C\xi} / n \right)$. Thus, (B.30) may be rewritten as

$$\sum_i^{(j)} \frac{x_{ij} - \mathbb{E} x_{ij}}{\sum_k^{(i,j)} x_{ik}} + \mathcal{O} \left(\frac{(\log n)^{2C+2C\xi}}{\sqrt{n}} \right), \quad (\text{B.33})$$

since the other error terms are dominated by the remaining one.

Note that x_{ij} does not appear in the summand's denominator and thus the denominator and numerator are independent. So we can use the large deviation estimate, Lemma E.4, with $a_i = x_{ij} - \mathbb{E} x_{ij}$ and $A_i^{-1} = \sum_k^{(i,j)} x_{ik}$. While the A_i are random, we may condition on their values and treat them as deterministic constants, then after we have used the LDE, we can bound them using the fact that we are on Ω . That is, on Ω

$$\left(\sum_k^{(i,j)} x_{ik} \right)^2 = (n-2)^2 (\mathbb{E} x_{ij})^2 + \mathcal{O} \left((\log n)^{2C+2C\xi} n \sqrt{n} \right) \quad (\text{B.34})$$

and so

$$\sum_k^{(j)} A_i^2 = \frac{1}{(n-2) (\mathbb{E} x_{ij})^2} + \mathcal{O} \left((\log n)^{2C+2C\xi} \frac{1}{n \sqrt{n}} \right). \quad (\text{B.35})$$

Thus the LDE gives us

$$\mathbb{P} \left[\left| \sum_i^{(j)} a_i A_i \right| \geq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \right] \leq e^{-\nu (\log n)^{1+\xi}}, \quad (\text{B.36})$$

which combined with (B.33)

$$\mathbb{P} \left[|W_j - 1| \geq \frac{C (\log n)^{2C+2C\xi}}{\sqrt{n}} \right] \leq e^{-\nu(\log n)^{1+\xi}}. \quad (\text{B.37})$$

The properties of high probability and the fact that we have n choices for j completes the proof. \blacksquare

Next we prove a lower bound on sums of edge weights, $w_I(S)$ and $w_O(S)$ for all $\emptyset \neq S \subsetneq V_n$. The proof relies on concentration inequalities for independent random variables and the simple fact that on Ω there is a constant $c > 0$ such that $w_{ij} \geq cn^{-1} \mathbb{1}(x_{ij} \geq c)$ for all $i, j \in V$.

LEMMA B.8. *On Ω , for all $\emptyset \neq S \subsetneq V_n$ and some small constant $c \equiv c_\mu > 0$, not dependent on n , we have the following bound*

$$|w_I(S)| = |w_O(S^c)| \geq c_\mu \min\{|S|, n - |S|\} \quad (\text{B.38})$$

with probability greater than

$$1 - e^{-\nu(\log n)^{1+\xi}}. \quad (\text{B.39})$$

PROOF. First note that as in the proof of Lemma B.7, we can argue that on Ω the sum $\sum_k^{(i)} x_{ik} \leq Cn$, see (B.29) for all $1 \leq i \leq n$. Moreover, by assumption on the distribution μ , we have $\mathbb{P}[x_{ij} > 0] = p > 0$ and thus there is a constant $c > 0$ such that $\mathbb{P}[x_{ij} \geq c] = p' > 0$. Therefore, on Ω

$$w_{ij} \geq cn^{-1} \mathbb{1}(x_{ij} \geq c). \quad (\text{B.40})$$

However, for each $1 \leq i, j \leq n$ with $i \neq j$, define $\beta_{ij} := \mathbb{1}(x_{ij} \geq c)$ which are independent Bernoulli random variables such that

$$\mathbb{P}[\mathbb{1}(x_{ij} \geq c) = 1] = p' > 0, \quad (\text{B.41})$$

since the x_{ij} are independent.

Let $|S| = k$. By definition

$$w_I(S) = \sum_{i \notin S} \sum_{j \in S} w_{ij} = w_O(S^c). \quad (\text{B.42})$$

Note that no diagonal terms are in these sums. Using (B.41),

$$\sum_{i \in S} \sum_{i \notin S} w_{ij} \geq \frac{c}{n} \sum_{i \in S} \sum_{i \notin S} \beta_{ij}. \quad (\text{B.43})$$

Note that now this is a sum of $k(n-k)$ independent random variables. So for fixed $\emptyset \neq S \subsetneq V_n$, by the Chernoff bound, Lemma E.3, the event

$$A_S := \left\{ \sum_{i \in S} \sum_{i \notin S} \beta_{ij} \leq (1 - 1/2)p'k(n-k) \right\} \quad (\text{B.44})$$

has probability less than

$$\exp \left(-\frac{1}{8} p' k(n-k) \right). \quad (\text{B.45})$$

Thus, by the union bound

$$\begin{aligned}
\mathbb{P} \left[\bigcap_{\emptyset \neq S \subsetneq V_n} A_S^c \right] &= 1 - \mathbb{P} \left[\bigcup_{\emptyset \neq S \subsetneq V_n} A_S \right] \\
&\geq 1 - \sum_{\emptyset \neq S \subsetneq V_n} \mathbb{P}[A_S] \\
&= 1 - \sum_{k=1}^{n-1} \binom{n}{k} \mathbb{P}[A_S] \\
&= 1 - \sum_{k=1}^{\lfloor \log n \rfloor} \binom{n}{k} \mathbb{P}[A_S] - \sum_{k=\lfloor \log n \rfloor+1}^{n-\lfloor \log n \rfloor-1} \binom{n}{k} \mathbb{P}[A_S] - \sum_{k=n-\lfloor \log n \rfloor}^{n-1} \binom{n}{k} \mathbb{P}[A_S] \\
&\geq 1 - 2n^{\log n} \exp(-p'n/8) - 2^n \exp(-p'n(\log n)/8) \\
&\geq 1 - 2 \exp(-(p'n/8 - (\log n)^2)) - \exp(-(p' \log n/8 - \log 2)n) \\
&\geq 1 - \exp(-cp'n),
\end{aligned} \tag{B.46}$$

for some $c > 0$. Finally, note

$$\frac{cp'}{2n} k(n-k) \geq \frac{cp'}{2} \min\{|S|, n-|S|\} \tag{B.47}$$

and

$$\exp(-cp'n) \leq e^{-\nu(\log n)^{1+\xi}} \tag{B.48}$$

for an appropriate choice of ξ and ν . ■

We now complete the proof of Theorem B.1 by putting together the results of Section A and the lemmata from this section.

PROOF OF THEOREM B.1. Again let $|S| = k$. We check that the assumptions of Theorem A.2 hold with high probability. Observe

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| = \left| \frac{w_O(S) - w_I(S)}{w_I(S)} \right| = \frac{|w_O(S) - w_I(S)|}{|w_I(S)|}. \tag{B.49}$$

Expanding the numerator, we get

$$w_O(S) - w_I(S) = \sum_{i \in S} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \in S} w_{ij} = \sum_{i \in S} \sum_{j \in V} w_{ij} - \sum_{i \in V} \sum_{j \in S} w_{ij} = \sum_{j \in S} (1 - W_j), \tag{B.50}$$

and similarly,

$$w_O(S) - w_I(S) = \sum_{i \in V} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \in V} w_{ij} = \sum_{j \notin S} (W_j - 1). \tag{B.51}$$

Thus Lemma B.7 implies

$$|w_O(S) - w_I(S)| \leq \min\{k, n-k\} \cdot \frac{C(\log n)^{C+C\xi}}{\sqrt{n}}, \tag{B.52}$$

for all S with high probability on Ω .

Lemma B.8 implies

$$|w_I(S)| \geq c \min\{|S|, n - |S|\} \quad (\text{B.53})$$

for all S with high probability on Ω . Putting this together we see

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \quad (\text{B.54})$$

for all S with high probability on Ω . However, by Lemma B.6, the event Ω holds with high probability itself and thus unconditionally (B.54) holds for all S with high probability.

Finally, applying Theorem A.2, we get

$$\sup_{r>0} |\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \quad (\text{B.55})$$

with high probability. ■

REMARK B.9 (On Theorem B.1). The parameter p , the probability that an edge of some weight exists between two directed vertices, can be interpreted as a measure of the sparseness of the population structure. We can ask how few interactions on average can individuals in a population have with others and still yield populations with Moran-type behavior? While p can be arbitrarily small, we have kept it constant and, in particular, not dependent on n . However, could p depend on n such that $p \rightarrow 0$ as $n \rightarrow \infty$ and still produce graphs which show Moran-type behavior? An obvious lower bound on the rate of p 's convergence to 0 is provided by the Erdős-Rényi model, which tells us that a graph is almost surely disconnected in the limit for $\sqrt{p} < (1 - \varepsilon)(1/n) \log n$ for any $\varepsilon > 0$. This bound follows by noting that $(1 - p)^2$ is the probability that there is no edge, in either direction, between two vertices and then applying the usual Erdős-Rényi threshold [18, 12]. There is much room between this lower bound and p constant—even whether such a sharp threshold for p exists is currently unclear. The issue is difficult to approach with naive simulations as the Moran process is not guaranteed to terminate on disconnected graphs.

APPENDIX C. GRAPHS WITHOUT OUTGOING WEIGHTS SUMMING TO 1

The Moran process on graphs may be generalized to no longer require the sum of the outgoing weights of each vertex to be 1. The process is still defined for a directed, weighted graph, $G \equiv G_n = (V_n, W_n)$, where $V \equiv V_n := \llbracket n \rrbracket$ but $W \equiv W_n := [w_{ij}]$ need not be a stochastic matrix (it must still have nonnegative entries). Instead of sampling *vertices* proportional to fitness and then choosing an outgoing edge with probability equal to its weight, we rather sample *edges* proportional to their weights and the fitness of the individual at the beginning of the edge. Once we select an edge the type of the individual at the end of the edge becomes the same as the individual at the beginning of the edge. It is easy to see that we get the original process as a special case of the new model. Again, the process X_t is Markovian with state space 2^V , where each X_t is the subset of V consisting of all the vertices occupied by mutants at time t . At time 0 a mutant is placed at one of the vertices uniformly at random. We then update as described above by choosing edges proportional to their weights and the individual's type at the edges origin. For example, the probability of choosing a particular edge (i, j) is

$$\frac{(r \mathbb{1}_{i \in X_t} + \mathbb{1}_{i \notin X_t}) w_{ij}}{r \sum_{i \in X_t} \sum_{j \in V} w_{ij} + \sum_{i \notin X_t} \sum_{j \in V} w_{ij}}. \quad (\text{C.1})$$

In this setting the robust isothermal theorem still holds.

THEOREM C.1 (ROBUST ISOTHERMAL THEOREM). *Fix $0 \leq \varepsilon < 1$. Let $G_n = (V_n, W_n)$ be a connected graph. If for all nonempty $S \subsetneq V_n$ we have*

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| \leq \varepsilon, \quad (\text{C.2})$$

then

$$\sup_{r>0} |\rho_{M_n}(r) - \rho_{G_n}(r)| \leq \varepsilon. \quad (\text{C.3})$$

PROOF. Just as before, we make the projection of the state space of all subsets of V , which records exactly which vertices are mutants, to the simpler state space $\{0, 1, \dots, n\}$, which records only the number of mutants. In the new model, we still have

$$\frac{p_+(S)}{p_-(S)} = r \frac{w_O(S)}{w_I(S)}. \quad (\text{C.4})$$

So the argument is exactly the same as the previous case. ■

APPENDIX D. MORE RANDOM GRAPHS

We can introduce a random graph model where the sum of the outgoing weights are not equal, that is, where individuals can contribute differentially to the next time point in a way not dependent on the genotype they carry. We change the model by not normalizing the outgoing edge weights. Following the Erdős-Rényi model, we produce a weighted, directed graph as follows: Consider an $n \times n$ matrix $X = [x_{ij}]$ with independent, identically distributed, nonnegative random variables for its entries. Since we do not need to normalize we take, $W = X$.

DEFINITION D.1 (GENERALIZED ERDŐS-RÉNYI RANDOM GRAPHS). *Let μ be a nonnegative distribution (not depending on n) with subexponential decay such that if $X \sim \mu$*

$$\mathbb{P}[X > 0] = p > 0 \text{ and } \mathbb{P}[X \geq x] \leq C e^{-x^c} \quad (\text{D.1})$$

for some positive constants p , c , and C and all $x > 0$. We denote the mean and standard deviation of X by μ_1 and σ respectively. We generate a family of random graphs $G_n = (V_n, X_n \equiv W_n)$ from μ , where x_{ij} are independent and distributed according to μ .

Now we have a similar theorem to before but the proof is easier.

THEOREM D.2. *Let $(G_n)_{n \geq 1}$ be a family of random graphs as in Definition D.1. Then there are constants $C > 0$ and $c > 0$, not dependent on n , such that the fixation probability of a randomly placed mutant of fitness $r > 0$ satisfies*

$$|\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \quad (\text{D.2})$$

uniformly in r with probability greater than

$$1 - \exp\left(-\nu (\log n)^{1+\xi}\right), \quad (\text{D.3})$$

for positive constants ξ and ν .

D.1. Proof of Theorem D.2. We now use the subexponential decay assumption to understand the typical behavior of the random variables x_{ij} .

DEFINITION D.3 (GOOD EVENTS Ω). *Let Ω be an n -dependent event such that the following hold:*

$$\Omega := \bigcap_{i,j=1}^n \left\{ x_{ij} \leq C (\log n)^C \right\} \cap \{G_n \text{ is connected}\}. \quad (\text{D.4})$$

LEMMA D.4. *The event Ω holds with high probability.*

PROOF. Identical to the previous proof. ■

Define the sum of the j th column as

$$W_j := \sum_{i=1}^n w_{ij} \quad (\text{D.5})$$

and the sum of the i th row as

$$\tilde{W}_i := \sum_{j=1}^n w_{ij}. \quad (\text{D.6})$$

LEMMA D.5. *On Ω , there are positive constants $c \equiv c_\mu$ and $C \equiv C_\mu$, not dependent on n , such that the following inequalities hold*

$$|W_j - \mu n| \leq C (\log n)^{C+C\xi} \sqrt{n} \quad (\text{D.7})$$

and

$$|\tilde{W}_i - \mu n| \leq C (\log n)^{C+C\xi} \sqrt{n} \quad (\text{D.8})$$

for all $i, j \in V$, with probability at least

$$1 - e^{-\nu(\log n)^{1+\xi}}. \quad (\text{D.9})$$

PROOF. Consider $a_i = w_{ij} - \mu$ or $a_i = w_{ji} - \mu$ and apply Lemma E.4. The claim follows immediately as there are $2n$ high probability events. ■

LEMMA D.6. *On Ω , for all $\emptyset \neq S \subsetneq V_n$ and some small constant $c \equiv c_\mu > 0$, not dependent on n , we have the following bound*

$$|w_I(S)| = |w_O(S^c)| \geq c_\mu |S| (n - |S|) \quad (\text{D.10})$$

with probability greater than

$$1 - e^{-\nu(\log n)^{1+\xi}}. \quad (\text{D.11})$$

PROOF. By assumption on the distribution μ , we have $\mathbb{P}[x_{ij} > 0] = p > 0$ and thus there is a constant $c > 0$ such that $\mathbb{P}[x_{ij} \geq c] = p' > 0$. Define $\beta_{ij} := \mathbb{1}(x_{ij} \geq c)$ which are independent Bernoulli random variables such that

$$\mathbb{P}[\beta_{ij} = 1] = p' > 0, \quad (\text{D.12})$$

since the x_{ij} are independent. Let $|S| = k$. By definition

$$w_I(S) = \sum_{i \notin S} \sum_{j \in S} x_{ij} = w_O(S^c). \quad (\text{D.13})$$

Using (B.41),

$$\sum_{i \in S} \sum_{i \notin S} x_{ij} \geq c \sum_{i \in S} \sum_{i \notin S} \beta_{ij}. \quad (\text{D.14})$$

Note that now this is a sum of $k(n-k)$ independent random variables. So for fixed $\emptyset \neq S \subsetneq V_n$, by the Chernoff bound, Lemma E.3, the event

$$A_S := \left\{ \sum_{i \in S} \sum_{i \notin S} \beta_{ij} \leq (1 - 1/2)p'k(n-k) \right\} \quad (\text{D.15})$$

has probability less than

$$\exp\left(-\frac{1}{8}p'k(n-k)\right). \quad (\text{D.16})$$

Thus, by the union bound

$$\begin{aligned} \mathbb{P}\left[\bigcap_{\emptyset \neq S \subsetneq V_n} A_S^c\right] &= 1 - \mathbb{P}\left[\bigcup_{\emptyset \neq S \subsetneq V_n} A_S\right] \\ &\geq 1 - \sum_{\emptyset \neq S \subsetneq V_n} \mathbb{P}[A_S] \\ &= 1 - \sum_{k=1}^{n-1} \binom{n}{k} \mathbb{P}[A_S] \\ &= 1 - \sum_{k=1}^{\lfloor \log n \rfloor} \binom{n}{k} \mathbb{P}[A_S] - \sum_{k=\lfloor \log n \rfloor+1}^{n-\lfloor \log n \rfloor-1} \binom{n}{k} \mathbb{P}[A_S] - \sum_{n-\lfloor \log n \rfloor}^{n-1} \binom{n}{k} \mathbb{P}[A_S] \\ &\geq 1 - 2n^{\log n} \exp(-p'n/8) - 2^n \exp(-p'n(\log n)/8) \\ &\geq 1 - 2 \exp(-(p'n/8 - (\log n)^2)) - \exp(-(p' \log n/8 - \log 2)n) \\ &\geq 1 - \exp(-cp'n), \end{aligned} \quad (\text{D.17})$$

for some $c > 0$. Finally, note

$$\exp(-cp'n) \leq e^{-\nu(\log n)^{1+\xi}} \quad (\text{D.18})$$

for an appropriate choice of ξ and ν . ■

We now complete the proof of Theorem D.2.

PROOF OF THEOREM D.2. Again let $|S| = k$. We check that the assumptions of Theorem C.1 hold with high probability. Observe

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| = \left| \frac{w_O(S) - w_I(S)}{w_I(S)} \right| = \frac{|w_O(S) - w_I(S)|}{|w_I(S)|}. \quad (\text{D.19})$$

Expanding the numerator, we get

$$w_O(S) - w_I(S) = \sum_{i \in S} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \in S} w_{ij} = \sum_{i \in S} \sum_{j \in V} w_{ij} - \sum_{i \in V} \sum_{j \in S} w_{ij} = \sum_{j \in S} (\tilde{W}_j - W_j), \quad (\text{D.20})$$

and similarly,

$$w_O(S) - w_I(S) = \sum_{i \in V} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \in V} w_{ij} = \sum_{j \notin S} (W_j - \tilde{W}_j). \quad (\text{D.21})$$

Thus Lemma D.5 implies

$$|w_O(S) - w_I(S)| \leq \min\{k, n - k\} \cdot C (\log n)^{C+C\xi} \sqrt{n}, \quad (\text{D.22})$$

for all S with high probability on Ω .

Lemma D.6 implies

$$|w_I(S)| \geq ck(n - k) \quad (\text{D.23})$$

for all S with high probability on Ω . Putting this together we see

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \quad (\text{D.24})$$

for all S with high probability on Ω . However, by Lemma D.4, the event Ω holds with high probability itself and thus unconditionally (B.54) holds for all S with high probability.

Finally, applying Theorem C.1, we get

$$\sup_{r>0} |\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \quad (\text{D.25})$$

with high probability. ■

APPENDIX E. LARGE DEVIATION ESTIMATES AND CONCENTRATION INEQUALITIES

In this section we provide a brief review of large deviation estimates and concentration inequalities with a focus on those used above. A large deviation estimate (LDE) controls atypical behavior of sums of independent (or sometimes weakly dependent) random variables, whereas a concentration inequality controls the convergence of an average of independent (or sometimes weakly dependent) random variables to their mean. For a more in-depth review of LDEs, see for example [12, 27]. Many LDEs follow directly by applying Markov's inequality, so we state this now.

THEOREM E.1 (MARKOV'S INEQUALITY). *Let X be a nonnegative random variable and $t > 0$. Then*

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}X}{t}. \quad (\text{E.1})$$

PROOF. Define the indicator random variable $\mathbb{1}_{X \geq t}$. Then $t\mathbb{1}_{X \geq t} \leq X$, thus $\mathbb{E}[t\mathbb{1}_{X \geq t}] \leq \mathbb{E}X$. Therefore,

$$\mathbb{P}[X \geq t] = \mathbb{E}\mathbb{1}_{X \geq t} \leq \frac{\mathbb{E}X}{t}. \quad \blacksquare$$

This very simple result has lots of scope. The general idea is to define a nonnegative, increasing function f of some random variable X and note that Markov's inequality implies

$$\mathbb{P}[X \geq t] \leq \mathbb{P}[f(X) \geq f(t)] \leq \frac{\mathbb{E}f(X)}{f(t)}. \quad (\text{E.2})$$

Normally, f is chosen as x^k or $e^{\lambda x}$ where k or $\lambda > 0$ is optimized to strengthen the inequality. If the random variable X is a sum of centered, independent random variables, $\sum_{i=1}^n (X_i - \mathbb{E}X_i)$, the function takes the form

$$\prod_{i=1}^n \exp(\lambda (X_i - \mathbb{E}X_i)). \quad (\text{E.3})$$

In this way we get several inequalities.

LEMMA E.2 (Hoeffding's inequality). *Suppose that X_1, \dots, X_n are i.i.d. Bernoulli random variables with parameter $p \in [0, 1]$. Define $X := \sum_{i=1}^n X_i$. Then*

$$\mathbb{P}[|X - \mathbb{E}X| \geq \delta\sqrt{n}] \leq 2 \exp(-2\delta^2) \quad (\text{E.4})$$

for all $\delta > 0$.

Lemma E.2 states that X fluctuates about its expectation on the order of \sqrt{n} , and the probability of a fluctuant greater than $\delta\sqrt{n}$ decays exponential with $\delta > 0$. The next lemma bounds fluctuation of larger orders, and thus they occur even more infrequently. We shall only need a lower bound in this case:

LEMMA E.3 (Multiplicative Chernoff bound). *Suppose that X_1, \dots, X_n are i.i.d. Bernoulli random variables with parameter $p \in [0, 1]$. Define $X := \sum_{i=1}^n X_i$. Then*

$$\mathbb{P}[X \leq (1 - \varepsilon)\mathbb{E}X] \leq \exp\left(-\frac{\varepsilon^2 p}{2}n\right) \quad (\text{E.5})$$

and

$$\mathbb{P}[X \geq (1 + \varepsilon)\mathbb{E}X] \leq \exp\left(-\frac{\varepsilon^2 p}{2}n\right). \quad (\text{E.6})$$

We remark that far more general statements of Lemmas E.2 and E.3 are possible, but we state only the versions we use in Section B.

Finally, we state a LDE for weighted sums of independent random variables with the following conditions on their moments:

$$\mathbb{E}X = 0, \quad \mathbb{E}|X|^2 = \sigma^2, \quad \text{and} \quad \mathbb{E}|X|^k \leq (Ck)^{Ck}, \quad (\text{E.7})$$

for some positive constant $C > 0$ (not dependent on n or k) and for $k \geq 1$.

LEMMA E.4. *Suppose the independent random variables $(a_i^{(n)})_{i=1}^n$ for $n \in \mathbb{N}$ satisfy (E.7) and that $(A_i^{(n)})_{i=1}^n$ for $n \in \mathbb{N}$ are constants in \mathbb{R} . Then*

$$\mathbb{P}\left[\left|\sum_{i=1}^n a_i A_i\right| \geq \sigma (\log n)^{C+C\xi} \left(\sum_{i=1}^n |A_i|^2\right)^{1/2}\right] \leq e^{-\nu(\log n)^{1+\xi}}. \quad (\text{E.8})$$

In words, we can bound the sum $\sum_{i=1}^n a_i A_i$ on the same order as the norm of the coefficients with high probability.

To prove this lemma we use a high-moment Markov inequality, so first we need a result bounding the higher moments of this sum.

LEMMA E.5. *Suppose the independent random variables $(a_i^{(n)})_{i=1}^n$ for $n \in \mathbb{N}$ satisfy (E.7) and that $(A_i^{(n)})_{i=1}^n$ for $n \in \mathbb{N}$ are constants in \mathbb{R} . Then*

$$\mathbb{E} \left| \sum_{i=1}^n a_i A_i \right|^k \leq (Ck)^{Ck} \left(\sum_{i=1}^n |A_i|^2 \right)^{k/2} \quad (\text{E.9})$$

PROOF. Without loss of generality let $\sigma = 1$. Let $A^2 := \sum_i |A_i|^2$, then by the classical Marcinkiewicz-Zygmund inequality [41] in the first line, we get

$$\mathbb{E} \left| \sum_i a_i A_i \right|^k \leq (Ck)^{k/2} \mathbb{E} \left| \left(\sum_i |A_i|^2 |a_i|^2 \right)^{1/2} \right|^k \quad (\text{E.10})$$

$$= (Ck)^{k/2} A^k \mathbb{E} \left[\left(\sum_i \frac{|A_i|^2}{A^2} |a_i|^2 \right)^{k/2} \right] \quad (\text{E.11})$$

$$\leq (Ck)^{k/2} A^k \mathbb{E} \left[\sum_i \frac{|A_i|^2}{A^2} |a_i|^k \right] \quad (\text{E.12})$$

$$= (Ck)^{k/2} A^k \sum_i \frac{|A_i|^2}{A^2} \mathbb{E} |a_i|^k \quad (\text{E.13})$$

$$\leq (Ck)^{Ck+k/2} A^k \quad (\text{E.14})$$

$$\leq (Ck)^{Ck} A^k, \quad (\text{E.15})$$

where we have used Jensen's inequality in the third line and assumption (E.7) in line 5. \blacksquare

PROOF OF LEMMA E.4. Without loss of generality let $\sigma = 1$. The proof is a simple application of Markov's inequality, Theorem E.1. Let $k = \nu (\log n)^{1+\xi}$, then by Lemma E.5, we get

$$\mathbb{P} \left[\left| \sum_i a_i A_i \right| \geq (\log n)^{C+C\xi} \left(\sum_i |A_i|^2 \right)^{1/2} \right] = \mathbb{P} \left[\left| \sum_i a_i A_i \right|^k \geq (\log n)^{Ck+Ck\xi} \left(\sum_i |A_i|^2 \right)^{k/2} \right] \quad (\text{E.16})$$

$$\leq \frac{\mathbb{E} |\sum_i a_i A_i|^k}{(\log n)^{Ck+Ck\xi} \left(\sum_i |A_i|^2 \right)^{k/2}} \quad (\text{E.17})$$

$$\leq \left(\frac{Ck}{(\log n)^{1+\xi}} \right)^{Ck} \quad (\text{E.18})$$

$$= (C\nu)^{C\nu(\log n)^{1+\xi}} \quad (\text{E.19})$$

$$\leq e^{-\nu(\log n)^{1+\xi}}, \quad (\text{E.20})$$

for $\nu \leq e^{-1}$ small enough. \blacksquare

REFERENCES

- [1] Benjamin Allen, Jeff Gore, and Martin A Nowak. Spatial dilemmas of diffusible public goods. *Elife*, 2:e01169, 2013.
- [2] Tibor Antal, Sidney Redner, and Vishal Sood. Evolutionary dynamics on degree-heterogeneous graphs. *Phys. Rev. Lett.*, 96:188104, 2006.
- [3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science.*, 286(5439):509–512, 1999.
- [4] Valmir C. Barbosa, Raul Donangelo, and Sergio R. Souza. Early appraisal of the fixation probability in directed networks. *Phys. Rev. E*, 82:046114, 2010.
- [5] Alain Barrat and Martin Weigt. On the properties of small-world network models. *Eur. Phys. J. B*, 13(3):547–560, 2000.
- [6] NH Barton. The probability of fixation of a favoured allele in a subdivided population. *Genet. Res.*, pages 149–157, 1993.
- [7] Alexei Borodin and Vadim Gorin. Lectures on integrable probability. *pre-print*, 2012.
- [8] Mark Broom and Jan Rychtář. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 464(2098):2609–2627, 2008.
- [9] Mark Broom and Jan Rychtář. *Game-theoretical models in biology*. Chapman & Hall/CRC mathematical and computational biology series. Chapman and Hall/CRC Press, Taylor and Francis Group, Boca Raton, FL, 2013.
- [10] Mark Broom, Jan Rychtář, and B Stadler. Evolutionary dynamics on small-order graphs. *J. Interdiscip. Math.*, 12:129–140, 2009.
- [11] Yu-Ting Chen. Sharp benefit-to-cost rules for the evolution of cooperation on regular graphs. *Ann. Appl. Probab.*, 23(2):637–664, 2013.
- [12] Fan Rong King Chung and Linyuan Lu. *Complex graphs and networks*, volume no. 107 of *CBMS regional conference series in mathematics* ;. American Mathematical Society, Providence, RI, 2006.
- [13] Percy Deift. Universality for mathematical and physical systems. *Proc. Int. Congr. Math.*, 2006.
- [14] Josep Díaz, Leslie Ann Goldberg, George B. Mertzios, David Richerby, Maria Serna, J., and Paul G. Spirakis. On the fixation probability of superstars. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 469:20130193, 2013.
- [15] Richard Durrett and Simon Levin. The Importance of Being Discrete (and Spatial). *Theor. Popul. Biol.*, 46:363–394, 1994.
- [16] Richard Durrett and Simon A Levin. Stochastic spatial models: a user’s guide to ecological applications. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.*, 343(1305):329–350, 1994.
- [17] László Erdős and Horng-Tzer Yau. Universality of local spectral statistics of random matrices. *Bull. Am. Math. Soc.*, 49(3):377–414, 2012.

- [18] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungarian Acad. Sci.*, 5, 1960.
- [19] Marcus Frean, Paul B Rainey, and Arne Traulsen. The effect of population structure on the rate of evolution. *Proc. R. Soc. B Biol. Sci.*, 280(1762):20130211, 2013.
- [20] Christoforos Hadjichrysanthou, Mark Broom, and Istvan Z Kiss. Approximating evolutionary dynamics on networks using a Neighbourhood Configuration model. *J. Theor. Biol.*, 312:13–21, 2012.
- [21] Michael P Hassell, Hugh N Comins, and Robert M May. Species coexistence and self-organizing spatial dynamics. *Nature*, 370:290–292, 1994.
- [22] Jaroslav Ispolatov and Michael Doebeli. Diversification along environmental gradients in spatially structured populations. *Evol. Ecol. Res.*, 11:295–304, 2009.
- [23] Alastair Jamieson-Lane and Christoph Hauert. Fixation probabilities on superstars, revisited and revised. *pre-print*, 2013.
- [24] Karin Johst, Michael Doebeli, and Roland Brandl. Evolution of complex dynamics in spatially structured populations. *Proc. R. Soc. B Biol. Sci.*, 266(1424):1147–1154, 1999.
- [25] Benjamin Kerr, Margaret a Riley, Marcus W Feldman, and Brendan J M Bohannan. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature*, 418(6894):171–4, July 2002.
- [26] Mickael Le Gac and Michael Doebeli. Environmental viscosity does not affect the evolution of cooperation during experimental evolution of colicigenic bacteria. *Evolution*, 64(2):522–33, 2010.
- [27] Michel Ledoux. *The concentration of measure phenomenon*, volume v. 89 of *Mathematical surveys and monographs*., American Mathematical Society, Providence, R.I., 2001.
- [28] Simon A Levin. Population dynamic models in heterogeneous environments. *Annu. Rev. Ecol. Syst.*, 7(1):287–310, 1976.
- [29] Simon A Levin and Robert T Paine. Disturbance, patch formation, and community structure. *Proc. Natl. Acad. Sci.*, 71(7):2744–2747, 1974.
- [30] Erez Lieberman, Christoph Hauert, and Martin A Nowak. Evolutionary dynamics on graphs. *Nature*, (JANUARY):312–316, 2005.
- [31] Takeo Maruyama. A Markov process of gene frequency change in a geographically structured population. *Genetics*, 76(2):367–377, 1974.
- [32] Takeo Maruyama. A simple proof that certain quantities are independent of the geographical structure of population. *Theor. Popul. Biol.*, 5(2):148–154, 1974.
- [33] Patrick Alfred Pierce Moran. *The statistical processes of evolutionary theory*. Oxford University Press, Oxford, England, 1962.
- [34] Mayuko Nakamaru and Simon A Levin. Spread of two linked social norms on complex interaction networks. *J. Theor. Biol.*, 230(1):57–64, 2004.
- [35] Martin A Nowak. *Evolutionary dynamics : exploring the equations of life*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, 2006.

- [36] Martin A Nowak, Franziska Michor, and Yoh Iwasa. The linear process of somatic evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 100(25):14966–14969, 2003.
- [37] Hisashi Ohtsuki, Christoph Hauert, Erez Lieberman, and Martin A Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441:502–505, 2006.
- [38] Z Patwa and Lindi M Wahl. The fixation probability of beneficial mutations. *J. R. Soc. Interface*, 5(28):1279–1289, 2008.
- [39] Paul B Rainey and Katrina Rainey. Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, 425:72–74, 2003.
- [40] B Sinervo and C M Lively. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature*, 380:240 – 243, 1996.
- [41] Daniel W Stroock. *Probability theory : an analytic view*. Cambridge University Press, Cambridge, UK, 2nd edition, 2011.
- [42] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
- [43] MC Whitlock. Fixation probability and time in subdivided populations. *Genetics*, 779(June):767–779, 2003.